

Н. С. Волохин¹✉, Р. В. Гришин¹

Разработка интеллектуальной системы распознавания геопозиции фотоматериалов

¹Сибирский государственный университет геосистем и технологий,
г. Новосибирск, Российская Федерация
e-mail: tesla.b2017@yandex.ru

Аннотация. В данной статье рассматривается задача автоматического определения географического положения по цифровым фотографиям. Это является одним из актуальных направлений на стыке компьютерного зрения и геоинформационных технологий. С ростом объема визуального контента в сети появляется все большая потребность в интеллектуальных решениях, способных оперативно и с высокой точностью определять место съемки изображения. Цель исследования заключалась в разработке системы геолокации снимков с использованием современных подходов машинного обучения. При этом акцент делался на микросервисной архитектуре и интеграции с популярными картографическими платформами. В результате был создан инструмент, который определяет геопозицию по фотографии и отображает объект на карте. Для реализации проекта были применены такие инструменты, как фреймворки Flask и TensorFlow, а также API внешних картографических сервисов. В ходе тестирования система показала высокую точность, устойчивость к изображениям с шумами и низким качеством. Разработанная система является стабильным, функциональным и удобным решением, позволяющим определять местоположение объектов на фотографиях. Архитектура системы позволяет легко масштабировать и улучшать ее, добавляя новые функции и совершенствуя алгоритмы распознавания.

Ключевые слова: геопозиционирование, компьютерное зрение, нейронные сети, машинное обучение, распознавание изображений, геоинформационные системы

N. S. Volokhin¹✉, R. V. Grishin¹

Development of an intelligent system for recognizing the geolocation of photomaterials

¹Siberian State University of Geosystems and Technologies, Novosibirsk, Russian Federation
e-mail: tesla.b2017@yandex.ru

Abstract. This article addresses the task of automatically determining geographic location from digital photographs. This is one of the relevant areas at the intersection of computer vision and geoinformation technologies. With the increasing volume of visual content on the internet, there is a growing need for intelligent solutions capable of quickly and accurately identifying the location where an image was taken. The aim of the research was to develop a geolocation system for images using modern machine learning approaches. The focus was on microservice architecture and integration with popular mapping platforms. As a result, a tool was created that determines the geolocation from a photograph and displays the object on a map. The project utilized tools such as the Flask and TensorFlow frameworks, as well as APIs from external mapping services. During testing, the system demonstrated high accuracy and resilience to images with noise and low quality. The developed system is a stable, functional, and user-friendly solution that allows for the determination of the location of objects in photographs. The architecture of the system enables easy scaling and improvement by adding new features and enhancing recognition algorithms.

Keywords: geo-positioning, computer vision, neural networks, machine learning, image recognition, geographic information systems

Введение

В последние годы с развитием технологий машинного обучения и компьютерного зрения значительный прогресс был достигнут в области распознавания объектов на изображениях и видео. Одним из наиболее актуальных направлений является использование этих технологий для распознавания геопозиции объектов по фотографиям. Система, способная автоматически определять местоположение изображенной на фотографии достопримечательности, открывает широкие возможности для туристических и географических приложений, а также для автоматизации процессов в области культурного наследия.

Актуальность разработки интеллектуальных систем распознавания геопозиции фотоматериалов становится особенно очевидной с учетом роста популярности платформ для путешественников, образовательных ресурсов и сервисов, связанных с изучением географических объектов. Одной из проблем таких сервисов является необходимость точного определения местоположения достопримечательностей, а также генерация контента, который мог бы удовлетворить интерес пользователей, обеспечивая информативность и актуальность.

Целью данной работы является разработка интеллектуальной системы для распознавания геопозиции фотоматериалов с использованием технологий машинного обучения. Система позволяет пользователям загружать фотографии достопримечательностей и получать информацию о местоположении объекта, включая страну, город, координаты, а также отображение на карте. Кроме того, система будет генерировать описание объекта на основе искусственного интеллекта, что обеспечит дополнение к стандартной информации о месте.

Методы и материалы

Среди уже существующих автоматических систем распознавания объектов на фотографиях и геопозиционирования можно отдельно упомянуть Google Lens и Mapillary. Google Lens проявляет собой высокую точность распознавания [1], однако его закрытая архитектура ограничивает доступ к исходному коду и возможности модификации системы. Mapillary, в свою очередь, ориентирован на картографирование, но требует ручной маркировки данных, что снижает степень автоматизации [2]. В отличие от указанных решений, предлагаемая система обладает рядом преимуществ: открытым исходным кодом, поддержкой пользовательских моделей машинного обучения и встроенной интеграцией переводчика и генерации текстовых описаний [3].

Разработанная система реализована в виде микросервисной архитектуры, что обеспечивает ее модульность, гибкость и хорошую масштабируемость. В состав решения входит сервис загрузки изображений, поддерживающий форматы JPEG, PNG и BMP. После загрузки изображение автоматически направляется в сервис распознавания, в основе которого лежит модель InceptionV3, дообученная

на наборе данных, включающем изображения достопримечательностей [4]. На этапе предобработки выполняется нормализация пикселей, а также применяется аугментация – повороты изображения и изменение яркости.

Распознанное название объекта передается в сервис, который с помощью API Nominatim преобразует его в координаты широты и долготы [5]. Для повышения устойчивости системы реализованы механизмы повторных запросов и кэширования полученных результатов. Финальным этапом является генерация текстовой аннотации, формируемой языковой моделью GigaChat.

В качестве базовой архитектуры выбрана модель. Обучение проводилось на 26 000 изображениях из набора Google Landmarks Dataset, с применением методов аугментации, что позволило существенно снизить риск переобучения [6].

Результаты

Для обработки изображений перед передачей в нейросетевую модель использовалась библиотека OpenCV [7]. Она обеспечила предварительную обработку загруженных файлов, включая изменение размера до 224 x 224 пикселей [8], нормализацию значений пикселей и конвертацию в цветовое пространство RGB [9]. Эти операции необходимы для корректной работы TensorFlow, так как многие предобученные модели ожидают изображения в определенном формате. OpenCV также использовался для проверки типа загруженного файла, устранения возможных артефактов и приведения изображений к единому стандарту [10], что улучшило качество работы системы [11].

Важным компонентом системы является модуль машинного обучения, который был реализован с использованием TensorFlow [12]. В качестве основы была выбрана предобученная модель InceptionV3, загруженная через TensorFlow Hub. Это позволило существенно сократить время обучения и достичь высокой точности распознавания [13]. Метод transfer learning использовался для адаптации модели к задаче определения достопримечательностей, что обеспечило возможность работы с реальными изображениями, загруженными пользователями [14]. TensorFlow также позволил реализовать возможность дальнейшего дообучения модели, что делает систему легко расширяемой [15]. Модель принимает обработанное изображение, анализирует его и возвращает идентификатор класса, соответствующий определенной достопримечательности [16].

На представленном графике показана динамика изменения точности модели в процессе обучения (рис. 1). По горизонтальной оси отложено количество эпох (Epoch), по вертикальной – значение точности (Accuracy). Синяя линия отражает точность на обучающей выборке (train), оранжевая – на валидационной (validation).

В начальный период обучения (1–3 эпохи) наблюдается резкий рост точности на обеих выборках. Интересно, что в первые эпохи валидационная точность оказывается выше обучающей.

На среднем этапе (4–10 эпохи) обе кривые демонстрируют стабильный рост. При этом значения точности на обучающей и валидационной выборках почти совпадают, что свидетельствует о хорошей способности модели к обобщению и отсутствии признаков переобучения.

В заключительной фазе (11–14 эпохи) рост точности продолжается, но становится не таким выраженным. Наблюдается небольшое превышение обучающей точности над валидационной, это нам говорит, что модель начала переобучаться, однако разрыв остается незначительным.

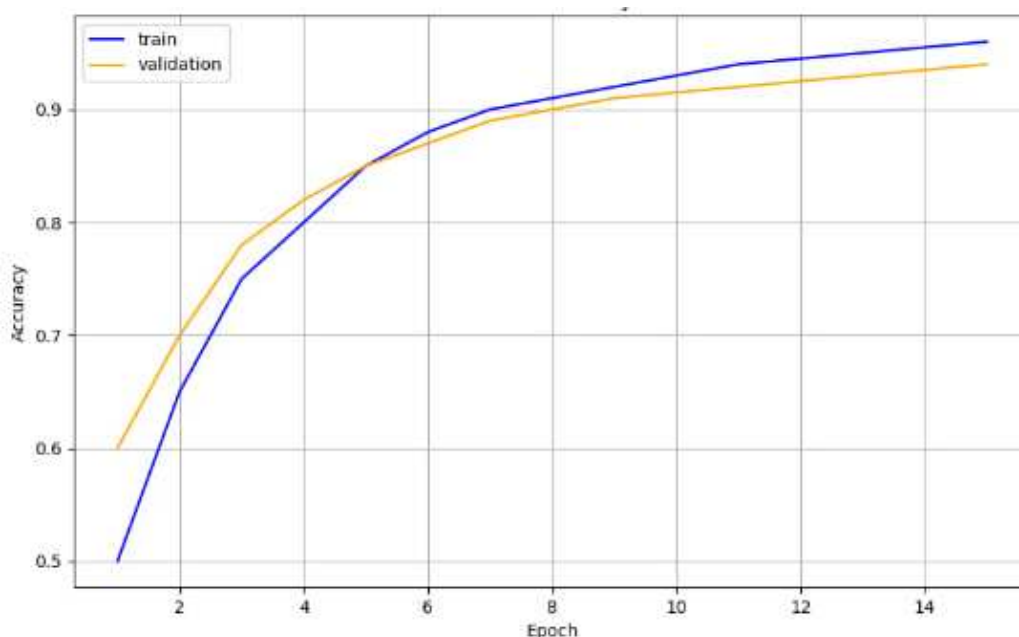


Рис. 1. Динамика изменения точности модели в процессе обучения

Разработанная система была интегрирована с веб-интерфейсом, для реализации были использованы HTML, CSS и JavaScript, Flask [17] на серверной стороне для обработки запросов, а также Bootstrap для удобной верстки [18]. На первой странице реализована интерактивная зона загрузки [19], поддерживающая drag-and-drop [20] и стандартный выбор файла через проводник (рис. 2).

После загрузки изображения и нажатия кнопки поиска система автоматически обрабатывает изображение и переходит на страницу результатов, которая содержит информацию об объекте: страну, город, название достопримечательности и краткое описание (рис. 3). Для отображения карты был использован Leaflet.js, интегрированный с OpenStreetMap [21]. На карте устанавливается метка с названием объекта, а при нажатии на нее появляется всплывающее окно с кратким описанием (рис. 4).

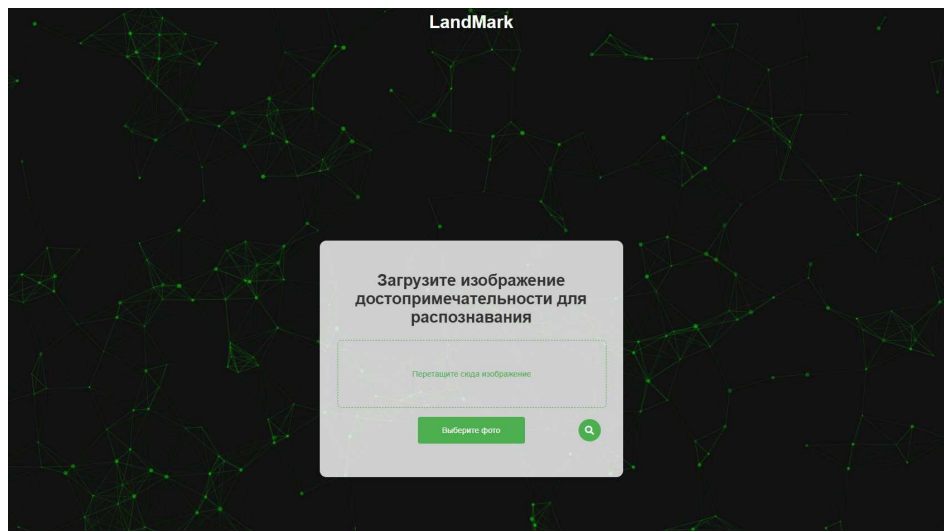


Рис. 2. Форма загрузки изображений

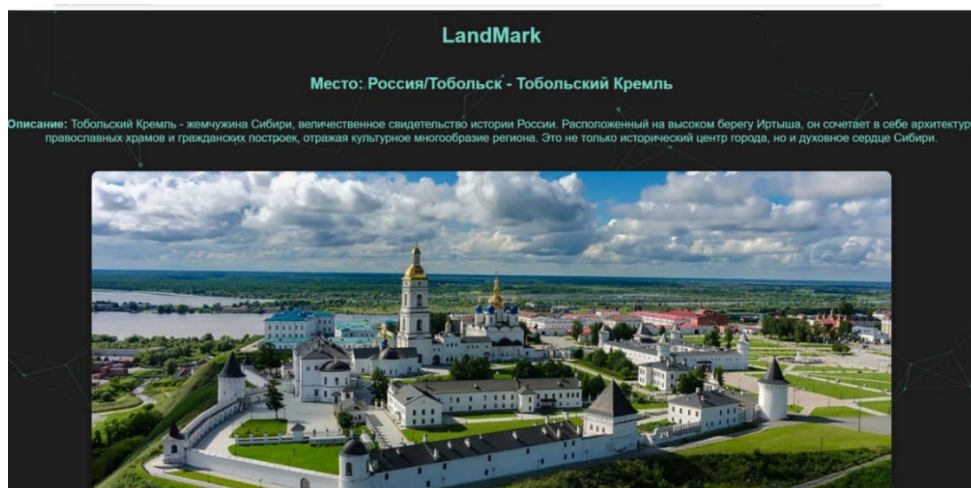


Рис. 3. Страница с распознанным объектом и его описанием



Рис. 4. Карта с меткой, координатами и описанием

Заключение

В ходе работы была разработана система определения геопозиций по фото-материалам, которая сочетает в себе возможности машинного обучения, обработки изображений, автоматического перевода и геолокации. Система позволяет пользователям загружать фотографии достопримечательностей и получать информацию о них, включая название, описание, координаты и местоположение на карте.

Анализ предметной области показал, что определение геопозиции по изображению является актуальной задачей, особенно в сфере туризма, культурного наследия и автоматизированных геоинформационных систем. Были сформулированы требования к системе, которые определили ее функциональные возможности и ключевые особенности, включая поддержку различных форматов изображений, автоматическое распознавание объектов, интеграцию с картографическими сервисами и удобный веб-интерфейс.

Разработанная система является стабильным, функциональным и удобным решением, позволяющим определять геопозицию объектов на фотографиях. Архитектура системы позволяет легко масштабировать и улучшать ее, добавляя новые функции и совершенствуя алгоритмы распознавания. В дальнейшем возможны дополнительные улучшения, включая расширение базы данных достопримечательностей, внедрение более точных моделей машинного обучения и интеграцию с дополнительными картографическими сервисами.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Particles.js: A lightweight JavaScript library for creating particles. – URL: <https://vincentgarreau.com/particles.js/> (дата обращения: 01.02.2025).
2. Вейланд Т., Мартинс Р., Тобиа М. PlaNet: геолокация фотографий с помощью сверточных нейронных сетей. – URL: <https://research.google.com/pubs/archive/45488.pdf> (дата обращения: 15.05.2025).
3. Dahua Technology. Продукты и решения. – URL: (дата обращения: 01.02.2025).
4. W3C. HTML5 Specification. – URL: <https://www.w3.org/TR/html5/> (дата обращения: 01.02.2025).
5. Гришин, Р. В. Способы извлечения метаданных о пространственном положении из хранилищ фотографий / Р. В. Гришин, А. А. Колесников. – Текст: непосредственный // Вестник СГУГиТ. – 2023. – Т. 28, № 6 – С. 57–66. – DOI 10.33764/2411-1759-2023-28-6-57-66.
6. Лю И., Дин Дж., Дэн Г. и др. Геолокация изображений с использованием больших моделей визуально-языкового понимания. – URL: <https://arxiv.org/abs/2408.09474> (дата обращения: 18.12.2024).
7. OpenCV Documentation. – URL: <https://docs.opencv.org/4.x/> (дата обращения: 01.02.2025).
8. Ху С., Фэн М., Нгуен Р., Ли Г. «CVM-Net: сеть сопоставления изображений разных ракурсов для геолокации». – URL: https://openaccess.thecvf.com/content_cvpr_2018/papers/Hu_CVM-Net_Cross-View_Matching_CVPR_2018_paper.pdf (дата обращения: 09.05.2025).
9. Keras Documentation. . – URL: <https://keras.io/guides/> (дата обращения: 01.02.2025).
10. Во Н., Хейс Дж. «Локализация и ориентация панорам улиц с помощью аэрофото-снимков». – URL: <https://arxiv.org/abs/1608.00161> (дата обращения: 10.05.2025).

11. Чжай М., Бессинджер З., Уоркман С., Джейкобс Н. «с». – URL: https://openaccess.thecvf.com/content_cvpr_2017/papers/Zhai_Predicting_Ground-Level_Scene_CVPR_2017_paper.pdf (дата обращения: 15.05.2025).
12. TensorFlow Documentation. – URL: <https://www.tensorflow.org/guide> (дата обращения: 01.02.2025).
13. Хейс Дж., Эфрос А. А. IM2GPS: оценка географической информации по одному изображению. – URL: <https://graphics.cs.cmu.edu/projects/im2gps/im2gps.pdf> (дата обращения: 15.05.2025).
14. Симонс М., Браун А. Геолокация изображений на основе механизма внимания. – URL: <https://onlinelibrary.wiley.com/doi/10.1155/2022/7168451> (дата обращения: 03.05.2025).
15. Вейанд Т., Костриков И., Филбин Дж. PlaNet: геолокация фотографий с помощью сверточных нейронных сетей. – URL: <https://research.google.com/pubs/archive/45488.pdf> (дата обращения: 15.05.2025).
16. Flask Documentation. – URL: <https://flask.palletsprojects.com/en/2.0.x/> (дата обращения: 01.02.2025).
17. Bootstrap Documentation. – URL: <https://getbootstrap.com/docs/5.0/getting-started/introduction/> (дата обращения: 01.02.2025).
18. Mozilla Developer Network. CSS: Cascading Style Sheets. – URL: <https://developer.mozilla.org/en-US/docs/Web/CSS> (дата обращения: 01.02.2025).
19. Leaflet.js: An Open-Source JavaScript Library for Mobile-Friendly Interactive Maps. – URL: <https://leafletjs.com/> (дата обращения: 01.02.2025).
20. Доу Ч., Ван Ц., Хан С. и др. GaGA: к интерактивному глобальному помощнику по геолокации. – URL: <https://arxiv.org/abs/2412.08907> (дата обращения: 12.12.2024).
21. Тейнер Й., Мюллер-Будак Э., Эверт Р. Интерпретируемая семантическая геолокация фотографий. – URL: <https://arxiv.org/abs/2104.14995> (дата обращения: 21.04.2025).

© Н. С. Волохин, Р. В. Гришин, 2025